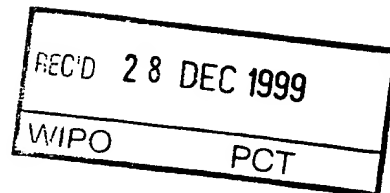# PRV ●

PATENT- OCH REGISTRERINGSVERKET
Patentavdelningen

SE 99 /1784

Intyg
Certificate

Härmed intygas att bifogade kopior överensstämmer med de handlingar som ursprungligen ingivits till Patent- och registreringsverket i nedannämnda ansökan.

This is to certify that the annexed is a true copy of the documents as originally filed with the Patent- and Registration Office in connection with the following patent application.

(71) *Sökande*   Karolinska Innovations AB, Stockholm SE
Applicant (s)

(21) *Patentansökningsnummer*   9803393-9
Patent application number

(86) *Ingivningsdatum*   1998-10-06
Date of filing

Stockholm, 1999-12-17

För Patent- och registreringsverket
For the Patent- and Registration Office

Anita Södervall

Avgift
Fee

PRIORITY
DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

A NOVEL COMPONENT IN THE HEDGEHOG SIGNALLING PATHWAY

Technical field

The present invention relates to novel molecules, such as proteins, polypeptides and nucleotides, involved in the hedgehog signalling pathway with putative involvement in embryonic development and carcinogenesis. The invention also relates to various novel advantageous uses of the molecules according to the invention in diagnosis and therapy.

Background

In the study of the development of cells, fruit flies have extensively been used as a model, as they are less complex than mammalian cells.

Pattern formation takes place through a series of logical steps, reiterated many times during the development of an organism. Viewed from a broader evolutionary perspective, across species, the same sort of reiterative pattern formations are seen. The central dogma of pattern formation has been described (Lawrence and Struhl, 1996). Three interlocking and overlapping steps are defined. Firstly, positional information in the form of morphogen gradients allocate cells into non-overlapping sets, each set founding a compartment. Secondly, each of these compartments acquire a genetic address, as a result of the function of active "selector" genes, that specify cell fate within a compartment and also instruct cells and their descendents how to communicate with cells in neighboring compartments. The third step involves interactions between cells in adjacent compartments, initiating new morphogen gradients, which directly organize the pattern.

Taking these steps in greater detail, one finds the first step in patterning to be the definition of sets of cells in each primordium. Cells are allocated according to their positions with respect to both dorsoventral and anterior/posterior axes by morphogen gradients. Allocation of cells in the dorsoventral axis constitutes the germ layers, such as mesoderm or neurectoderm.

In segmentation, the second step (the specification of cell fate in each compartment) is carried out by the gene *engrailed* and elements of the bithorax complex. *Engrailed* defines anterior and posterior compartments both in segmentation and in limb specification.

5

The third step in pattern formation, secretion of morphogens, functions to differentiate patterns within compartments (and thereby establish segment polarity). Initially, all cells within a compartment are equipotent, but they become diversified to form pattern. Pattern formation depends on gradients of morphogens, gradients initiated along compartment boundaries. Such gradients are established by a short-range signal induced in all the cells of the compartment in which the above mentioned selector gene *engrailed* is active. For segment polarity, this signal is Hedgehog. In the adjacent compartment the selector gene is inactive, ensuring that the cells are sensitive to the signal. The Hedgehog signal range is probably only a few rows of cells wide; responding cells become a linear source of a long-range morphogen, that diffuses outward in all directions. There are three known Hedgehogs, Sonic (SHH), Indian (IHH) and Desert (DHH). The proteins they encode can substitute each for each other, but in wildtype animals, their distinct distributions result in unique activities. SHH controls the polarity of limb growth, directs the development of neurons in the ventral neural tube and patterns somites. IHH controls endochondral bone development and DHH is necessary for spermiogenesis. Vertebrate hedgehog genes are expressed in many other tissues, including the peripheral nervous system, brain, lung, liver, kidney, tooth primordia, genitalia and hindgut and foregut endoderm.

Thus, segment polarity genes have been identified in flies as mutations, which change the pattern of structures of the body segments. Mutations in these genes cause animals to develop the changed patterns on the surfaces of body segments, the changes affecting the pattern along the head to tail axis. For example, mutations in the gene *patched* cause each body segment to develop without the normal structures in the center of each segment. Instead there is a mirror image of the pattern normally found in the anterior segment. Thus, cells in the center of the segment make the

wrong structures, and point them in the wrong direction with reference to the over all head-to-tail polarity of the animal.

5 About sixteen genes in the class are known. The encoded proteins include kinases, transcription factors, a cell junction protein, two secreted proteins called wingless (WG) and the above mentioned Hedgehog (HH), a single transmembrane protein called patched (PTC) and some novel proteins not related to any known protein. All of these proteins are beleived to work together in signaling pathways that inform 10 cells about their neighbors in order to set cell fates and polarities.

PTC has been proposed as a receptor for HH protein based on genetic experiments in flies. A model for the relationship is that PTC acts through a largely unknown pathway to inactivate both its own transcription and the transcription of the *wingless* segment polarity gene. This model proposes that HH protein, secreted from adjacent 15 cells, binds to the PTC receptor, inactivates it and thereby prevents PTC from turning off its own transcription or that of *wingless*. A number of experiments have shown coordinate events between PTC and HH.

20 Human *patched* gene (PTCH) was recently identified as the gene responsible for the nevoid basal cell carcinoma syndrome (NBCCS), also known as the Gorlin Syndrome, which is an autosomal dominant disorder that predisposes to both cancer and developmental defects (Gorlin (1995) *Dermatologic Clinics* 13:113-125) characterized by multiple basal cell carcinomas (BCCs), medulloblastomas and ovarian fibromas as well as numerous developmental anomalities (Hahn, H., Wicking, C., 25 Zaphiropoulos, P.G., Gailani, M.R., Shanley, S., Chidambaram, A., Vorechovsky, I., Holmberg, E., Undén, A.B., Gillies, S., Negus, K., Smyth, I., Pressman, C., Leffell, D.J., Gerrard, B., Goldstein, A.M., Dean, M., Toftgård, R., Chenevix-Trench, G., Wainright, B. and Bale, A.E. (1996): "Mutations of the human homolog of Drosophila patched in the nevoid basal cell carcinoma syndrome", Cell **85**, 841-851; 30 and Johnson, R.L., Rothman, A.L., Xie, J., Goodrich, L.V., Bare, J.W., Bonifas, J.M., Quinn, A:G., Myers, R.M., Cox, D.R., Epstein, E.H. Jr and Scott, M.P.

(1996): "Human homolog of patched, a candidate gene for the basal cell nevus syndrome", Science **272**, 1668-1671). PTCH codes for a membrane receptor of the autolytically cleaved (protein spliced), amino terminal domain of *sonic hedgehog* (SHH) (Mariago, V., Davey, R.A., Zuo, Y., Cunningham, J.M. and Tabin, C.J.

5 (1996): "Biochemical evidence that patched is the Hedgehog receptor", Nature **384**, 176-179; and Stone, D.M., Hynes, M., Armanini, M., Swanson, T.A., Gu, Q., John-son, R.L., Scott, M.P., Pennica, D., Goddard, A., Phillips, H., Noll, M., Hooper, J.E., de Sauvage, F. and Rosenthal, A. (1996): "The tumor-suppressor gene patched encodes a candidate receptor for Sonic hedgehog", Nature **384**, 129-134). In the

10 non-signalling state, PTCH is thought to inhibit the consecutive signalling of anot-her membrane protein, *smoothened* (SMO), however binding of SHH to PTCH re-leives this inhibition (Goodrich, L.V., Milenkovic, L., Higgins, K.M. and Scott, M.P. (1997): "Altered neural cell fates and medullablastom in mouse patched mu-tants", Science **277**, 1109-1113). This cascade of signalling events, best characteri-

15 zed in Drosophila, also involves a number of intracellular components including *fu-sed* (a serine threonine kinase), *suppressor of fused*, *costal 2*, and *cubitus interrup-tus* (Ruiz i Altaba, A.,: "Catching a Gli-mpse of Hedgehog" (1997) Cell **90**, 193-196). The latter is a transcription factor that positively regulates the expression of target genes which also include PTCH itself.

20

Mutations in the PTCH gene have been identified in both sporadic and familial BCCs (Gailani, M.R., Ståhle-Bäckdahl, M., Leffell, D.J., Glynn, M., Zaphiropou-los, P.G., Pressman, C., Undén, A.B., Dean, M., Brash, D. E., Bale, A.E. and Toft-gård, R. (1996): "The role of human homologue of Drosophila patched in sporadic

25 basal cell carcinomas" Nature Genet. **14**, 78-81). The lack of the normal PTCH protein in these cells allows the constitutive signalling of SMO to occur, resulting in the accumulation of mutant PTCH mRNAs (Undén, B. A., Zaphiropolous, P.G., Bruce, K., Toftgård, R., and Ståhle-Bäckdahl, M. (1997): "Human patched (PTCH) mRNA is overexpressed consistently in tumor cells of both familial and sporadic

30 basal cell carcinoma", Cancer Res. **57**, 2336-2340).

WO 96/11260 discloses the isolation of *patched* genes and the use of the PTC protein to identify ligands, other than the established ligand Hedgehog, that bind thereto.

5

However, there is still a need of a further understanding of the SHH/PTCH cell signalling, which may be provided by disclosure of further genes, peptides and proteins involved therein.

10 Summary of the invention

The present invention provides a significant step forward regarding the understanding of the above described pathway. By a combination of cDNA library and RACE analysis a novel human *patched*-like gene (PTCH2) has been cloned and sequenced. Several alternatively spliced mRNA forms of PTCH2 have been ideintified,

15 including transcripts lacking segments thought to be involved in sonic hedgehog (SHH) binding and mRNAs with differentially defined 3' terminal exons. Accordingly, the invention relates to isolated such mRNAs as well as to cDNAs complementary thereto.

20 Brief description of the drawings

Figure 1 shows SEQ ID NO 1, which is the amino acid sequence encoded by the novel human *patched* 2 gene.

Figure 2 shows SEQ ID NO 2, which is the nucleotide sequence encoding the protein disclosed by SEQ ID NO 1.

Figure 3 shows SEQ ID NO 3, wherein exons and introns are designated in the genomic sequence of the novel human patched 2 gene.

30 Figure 4A discloses an amino acid sequence comparison of the human PTCH2 (upper lines) and PTCH1 (lower lines) sequences.

Figure 4B is a representation of the alternative splicing events that result in different C-termini.

5   Figure 4C is a representation of the different variations of spliced transcripts encompassing exon 1 and exon 2 sequences.

Figure 5A is a dark-field photomicrograph of a BCC tumor hybridized with ^{35}S-labeled antisense probe showing abundant signal for PTCH1 mRNA (light grains) in
10   all BCC tumor cells.

Figure 5B discloses PTCH2 mRNA overexpression in BCC and is in contrast mainly expressed in the basaloid cells in the periphery of the tumor nests.

15   Figure 5C is another BCC showing a strong PTCH2 mRNA signal in the periphery of the tumor nest (Tu), wheras no signal is detected in epidermis (Ep).

Figure 5D are sections of the same tumor (C) hybridized with the PTCH2 sense probe showed no signal.
20

Figure 5E shows immunoreactivity for Ki-67.

Figure 5F discloses how tumor nests under high power magnification demonstrate abundant PATCH2 mRNA signal (black grains) in the dark basaloid tumor cells and
25   lower signal in the center (arrow).

Definitions

The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid poly-
30   mers in which one or more amino acid residue is an artificial chemical analogue of a

corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers.

The terms "isolated" "purified" or "biologically pure" refer to material which is substantially or essentially free from components which normally accompany it as found in its native state.

The term "nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogs of natural nucleotides that can function in a similar manner as naturally occurring nucleotides.

A "label" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include ^{32}P, fluorescent dyes, electron-dense reagents, enzymes (*e.g.*, as sommonly used in a ELISA), biotin, dioxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available (*e.g.*, the peptide of SEQ ID NO 1 can be made detectable, *e.g.*, by incorporating a radio-label into the peptide, and used to detect antibodies specifically reactive with the peptide).

As used herein a "nucleic acid probe" is defined as a nucleic acid capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include natural (*i.e.* A, G, C, or T) or modified bases (7-deazaguanosine, inosine, *etc.*) In addition, the bases in a probe may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, for example, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather that phosphodiester linkages. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions. The probes

are preferably directly labeled as with isotopes, chromophores, lumiphore, chromogens, or indirectly labeled such as with biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the selct sequence or subsequence.

5

A "labeled nucleic acid probe" is a nucleic acid probe that is bound, either covalently, through a linker, or through ionic, van der Waals or hydrogen bonds to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe.

10

The term "target nucleic acid" refers to a nucleic acid (often derived from a biological sample), to which a nucelic acid probe is designed to specifically hybridize. It is either the presence or absence of the target nucleic acid that is to be detected, or the amount of the target nucleic acid that is to be quantified. The target nucleic acid has

15 a sequence that is complementary to the nucleic acid sequence of the corresponding probe directed to the target. The term target nucleic acid may refer to the specific subsequence of a larger nucleic acid to which the probe is directed or to the ovarall sequence (*e.g.*, gene or mRNA) whose expression level it is desired to detect. The difference in usage will be apparent from context.

20

The term "recombinant" when used with reference to a cell, or nucleic acid, or vector, indicates that the cell, or nucleic acid, or vector, has been modified by the introduction of a heterologous nucleic acid or the alteration of a native nucleic acid, or that the cell is derived from a cell so modified.

25

The term "identical" in the context of two nucleic acids or polypeptide sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorith of Smith and Waterman (1981) *Adv. Appl. Math.* 2: 482, by the homology alignment algorithm of Needleman and

30 Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity method of Pearson

and Lipman (1988) *Proc. Natl. Acad. Sci.* USA 85: 2444, by computerized implementations of these algorithms (GAP, GESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI) or by inspection. The BLAST algorithm performs a statistical analy-

5 sis of the similarity between two sequences; *see e.g.,* Karlin and Altschul (1993) Proc. *Nat'l Acad. Sci. USA* 90: 5873-5787.

The term "substantial identity" or "substantial similarity" in the context of a polypeptide indicates that a polypeptides comprises a sequence with at least 70%

10 sequence identity to a reference sequence, or preferably 80%, or more preferably 85% sequence identity to the reference sequence, or most preferably 90% identity over a comparison window of about 10-20 amino acid residues. An indication that two polypeptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a

15 polypeptide is substantially identical to a second polypeptide, for example, where the two peptides differ only by a conservative substitution.

An indication that two nucleic acid sequences are substantially identical is that the polypeptide which the first nucleic acid encodes is immunologically cross reactive

20 with the polypeptide encoded by the second nucleic acid. Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions.

The phrase "hybridizing specifically to", refers to the binding, duplexing, or hy-

25 bridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (*e.g.,* total cellular) DNA or RNA. The term "stringent conditions" refers to conditions under which a probe will hybridize to its target subsequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances.

30 Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point ™ for

the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. (As the target sequences are generally present in excess, at Tm, 50% of

5 the probes are occupies at equilibrium). Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for whort probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may

10 also be achieved with the addition of destabilizing agents such as formamide.

The term "antibody" refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen).

15

A "chimeric antibody" is an antibody molecule in which (a) the constant region, or a portion thereof, is altered, replaced or exchanged so that the antigen binding site (variable region) is linked to a constant region of a different or altered class, effector function and/or species, or an entirely different molecule which confers new prop-

20 erties to the chimeric antibody, e.g., an enzyme, toxin, hormone, growth factor, drug, etc.; or (b) the variable region, or a portion thereof, is altered, replaced or exchanged with a variable region having a different or altered antigen specificity.

The term "immunoassay" is an assay that utilizes an antibody to specifically bind an

25 analyte. The immunoassay is characterized by the use of specific binding properties of a particular antibody to isolate, target, and/or quantify the analyte.

The phrases "specifically binds to a protein" or "specifically immunoreactive with", when referring to an antibody refers to a binding reaction which is determinative of

30 the presence of the protein in the presence of a heterogeneous population of proteins

and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind preferentially to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to a protein under such conditions requires an antibody that is selected for its specificity for a particu-

5 lar protein. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antoibodies specifically immunoreactive with a protein. See harlow and Lane (1988) *Antibodies, A Laboratory Manual,* Cold Spring Harbour Publications, New York, for a description of immu-

10 noassay formats and conditions that can be used to determine specific immunoreactivity.

A "gene product", as used herein, refers to a nucleic acid whose presence, absence, quantity, or nucleic acid sequence is indicative of a presence, absence, quantity, or

15 nucleic acid composition of the gene. Gene products thus include, but are not limited to, and mRNA transcript acDNA reverse transcribed from an mRNA, and RNA transcribed from that cDNA, a DNA amplified from the cDNA, an RNA transcribed from the amplified DNA or subsequences of any of these nucleic acids. Polypeptides expressed by the gene or subsequences thereof are also gene products. The

20 particular type of gene product will be evident from the context of the usage of the term.

Detailed description of the invention

In a first aspect, the present invention relates to an isolated human protein, or an

25 analogue or a variant thereof, capable of participating in the human PTCH/SHH pathway during embryonic development and/or carcinogenesis, such as basal cell carcinoma. The novel protein according to the invention is encoded by a novel gene, which isolated nucleic acid is described in detail below and which is denoted *patched 2* (PTCH2) due to its similarities with *patched 1* (PTCH1). Accordingly, the

30 protein according to the invention exhibits substantial differences in sequence and functions when compared to human PTCH1protein. The protein according to the

invention is best characterized by its functions which when compared to human
PTCH1 are similar but distinct therefrom in certain ways, more specifically disclo-
sed below in the section "Results and discussion". The novel human PTCH2 protein
according to the invention is also distinct from the previously isolated mouse

5   PTCH2. Thus, in the preferred embodiment thereof, it comprises a substantial part
of the amino acid sequence disclosed in SEQ ID NO. 1 of Figure 1, even though it
is to be understood that the present invention encompasses any fragment, analogue
or variant thereof exhibiting the biological functions of the PTCH2 protein disclosed
herein. Thus, preferably, the present protein comprises at least about 1000, more

10   preferably at least about 1040 and most preferably essentially all of the amino acids
of the sequence denoted SEQ ID NO. 1 of Figure 1, such as about 1100.

The proteins according to the invention are easily prepared by someone skilled in
this field by recombinant DNA techniques using the molecules disclosed below or

15   any synthetic method (*see e.g.* Barany and Merrifield, Solid-Phase Peptide synthe-
sis, pp. 3-284 in The Peptides: Analysis, Synthesis, Biology, Vol. 2: Special Met-
hods in Peptide synthesis, Part A, Merrifield *et al.*, J. Am. Chem. Soc., 2149-2156).

The present invention also relates to the use of the peptides, polypeptides and pro-

20   teins disclosed herein as lead compounds in methods aimed at finding novel sub-
stances, such as substances exhibiting equivalent or even more advantageous pro-
perties than the lead compounds as such. The invention also relates to proteomic
methods wherein the present molecules are used as well as to such a use *per se*.

25   A second aspect of the present invention is a nucleic acid encoding a protein, an
analogue or a variant thereof as defined above, that is, the protein coding region of
the novel human isolated PTCH2 gene. The PTCH2 gene is 57% identical to
PTCH1 and 91% identical to the published mouse Ptch2 sequence (*see* Motoyama
et al., (1998), supra). Thus, preferably, the nucleic acid according to the present in-

30   vention comprises at least about 3000 bases, more preferably at least about 3094 ba-

ses and most preferably essentially all of the sequence denoted SEQ ID NO 2 of Figure 2.

In a specific aspect, the present invention relates to the isolated human genomic PTCH2 nucleic acid comprising parts or all of the sequence denoted SEQ ID NO 3 of Figure 3. This aspect of the invention advantageously enables design of suitable PCR primers, which in turn enables screening for mutations of all of the coding sections thereof, *e.g.* by SSCP analysis, sequencing, or any other suitable method known to someone skilled in this field. Thus, the novel human PTCH2 gene according to the invention has been localized by radiation hybrid mapping to chromosome 1p32-35 with D1S211 and WI-1404 as closest flanking markers and with an estimated localization 5.5cR from D1S443. This region is often lost by LOH in various different tumor types, such as neuroblastoma, melanoma, breast cancer, colon cancer *etc.* Accordingly, PTCH2 is a candidate for a tumor suppressor gene in this region and the present invention also encompass diagnostic methods based on this new disclosure.

To this chromosomal region, three cancer predisposition syndromes have also been mapped, namely, familial melanoma CMM1, modifier locus for familial adenomatous polyposis hMom1 and Michelin Tire Baby Syndrome. PTCH2 is further a candidate for the gene behind these heritary syndromes. The present molecules are therefore advantageously used in the context of these conditions, *e.g.* in therapy and/or diagnosis, such as in assays.

Further, the invention also relates to various PCR primers based on intronic sequences, allowing amplification of all coding sequence. Such primers are advantageously used for mutation screening.

Further, the present invention also relates to the any isolated nucleic acid capable of specifically hybridizing to a nucleic acid according to the invention. In addition, the invention also relates to such an isolated nucleic acid which comprises one o more

mutations compared to the genomic sequence as well as the use of the novel isolated nucleic acids, *e.g.* to identify mutations for diagnostic and/or therapeutic purposes.

Further embodiments of this aspect of the invention includes nucleic acid probes, *e.g.* DNA probes, labelled nucleic acids, cDNAs, RNAs *etc.*, that is, all gene products obtainable by someone skilled in this field based on the novel isolated human PTCH2 gene.

Another aspect of the invention is a nucleic acid corresponding to any one of the splicing variants disclosed in Figure 4B, a protein or polypeptide encoded thereof as well as various uses thereof.

As regards the preparation of nucleic acids according to the invention, any suitable recombinant DNA technique or synthetic method may be used. (For general laboratory procedures useful in this context, *see e.g.* Sambrook *et al.*, Molecular Cloning, A Laboratory Manual, 2nd ed., Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989; Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol. 153, Academic Press, Inc., San Diego, CA; Current Protocols in Molecular Biology, F.M. Ausbel *et al.*, eds., Current Protocols (1994)).

A further aspect of the present invention is a vector comprising a nucleic acid as defined above. Vectors are *e.g.* useful for transforming cells <u>in vitro</u> or <u>in vivo</u> to express the proteins and peptides according to the invention and may e.g be plasmids, viruses *etc.*

Another aspect of the invention is a recombinant cell, such as a eucaryotic, *e.g.* a mammalian cell, or a procaryotic cell, *e.g.* a bacteria, comprising a vector as defined above. Such cells may *e.g.* be used to monitor expression levels of the proteins and polypeptides according to the invention in a wide variety of contexts. For example, when the effects of a drug is to be determined, the drug will be administered to the

transformed organism, tissue or cell. Accordingly, model systems including such cells are another aspect of the invention.

A further aspect of the invention is an antibody, such as a monoclonal or polyclonal antibody, which specifically binds to a protein or polypeptide according to the invention. An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (V_L) and variable haeavy chain (V_H) refer to these light and heavy chains, respectively.

The invention also encompasses chimeric or other antibodies that binds the present proteins or polypeptides. Further, the invention also relates to the use of the present antibodies in assays. (In this context, *see e.g. Fundamental Immunology*, Third Edition, W.E. Paul, ed., Raven Press, N.Y. 1993).

Further, the invention also relates to a recombinant cell expressing an antibody according to the invention.

In general, prokaryotes can be used for cloning the DNA sequences encoding a human anti-PTCH2 immunoglobulin chain. *E. coli* is one prokaryotic host particularly useful for cloning the DNA sequences of the present invention. Microbes, such as yeast are also useful for expression. *Saccharomyces* is a preferred yeast host, with suitable vectors having expression control sequences, an origin of replication, termination sequences and the like as desired. Typical promoters include 3-phosphoglycerate kinase and other glycolytic enzymes. Inducible yeast promoters include, among others, promoters from alcohol dehydrogenase 2, isocytochrome C, and enzymes responsible for maltose and galactose utilization.

Mammalian cells are a particularly preferred host for expressing nucleotide seg-
ments encoding immunoglobulins or fragments thereof (*see, e.g.* Winnacker, *From
Genes to Clones,* VCH Publishers, N.Y., 1987). A number of suitable host cell lines
5 capable of secreting intact heterologous proteins have been developed in the art, and
include CHO cell lines, various COS cell lines, HeLa cells, L cells and myeloma
cell lines. Preferably, the cells are nonhuman. Expression vectors for these cells can
include expression control sequences, such as an origin of replication , a promoter,
an enhancer (Queen *et al.* (1986) *Immunol. Rev.* 89:49), and necessary processing
10 information sites, such as ribosome binding sites, RNA splice sites, polyadenylation
sites, and transcriptional terminator sequences. Preferred expression control se-
quences are promoters derived from endogenous genes, cytomegalovirus, SV40,
adenovirus, bovine pappillomavirus, and the like (*see, e.g.,* Co *et al.* (1992) *J. Im-
munol.* 1458: 1149).

15

An additional aspect of the present invention is a kit for the detection of a human
PTCH2 gene or polypeptide comprising in a container a molecule selected from the
group consisting of a nucleic acid, a polypeptide or a protein or an antibody accor-
ding to the invention. Further suitable components of such a kit are easily determi-
20 ned by someone skilled in this field as are the conditions for the use thereof.

Further, the invention also realtes to the use of a nucleic acid selected from the group consisting of SEQ ID NO. 2 and SEQ ID NO. 3 in gene therapy. For a review of gene therapy procedures, *see* Anderson, *Science* (1992) 256:808-813; Nabel and Felgner (1993) *TIBTECH* 11: 211-217; Mitani and Caskey (1993) *TIBTECH* 11:

5   162-166; Mulligan (1993) *Science* 926-932; Dillon (1993) *TIBTECH* 11: 167-175; Miller (1992) *Nature* 357: 455-460; Van Brunt (1988) *Biotechnology* 6(10): 1149-1154; Vigne (1995) *Restorative Neurology and Neuroscience* 8: 35-36; Kremer and Perricaudet (1995) *British Medical Bulletin* 51(1) 31-44; Haddada *et al.* (1995) in *Current Topics in Microbiology and Immunology* Doerfler and Böhm (eds)

10   Springer-Verlag, Heidelberg Germany; and Yu *et al.*, *Gene Theraphy*(1994) 1:13-26.

Delivery of the gene or genetic material into the cell is the first critical step in gene therapy treatment of disease. A large number of delivery methods are well known to

15   those of skill in the art. Such methods include, for example liposome-based gene delivery (Debs and Zhu (1993) WO 93/24640; Mannino and Gould-Fogerite (1988) *BioTechniques* 6(7): 682-691; Rose U.S. Pat No. 5,279,833; Brigham (1991) WO 91/06309; and Felgner *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84: 7413-7414), and replication-defective retroviral vectors harboring a therapeutic polynucleotide se-

20   quence as part of the retroviral genome (*see, e.g.,* Miller *et al.* (1990) *Mol. Cell. Biol.* 10:4239 (1990; Kolberg (1992) *J. NIH Res.* 4:43, and Cornetta *et al. Hum. Gene Ther.* 2:215 (1991)). Widely used retroviral vectors include those based upon nurine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations

25   thereof. *See, e.g.,* Buchscher *et al.* (1992) *J. Virol.* 66 (5) 2731-2739; Johann *et al.* (1992) *J. Virol.* 66 (5):1635-1640 (1992); Sommerfelt *et al.,* (1990) *Virol.* 176:58-59; Wilson *et al.* (1989) *J. Virol.* 63:2374-2378; Miller *et al., J Virol.* 65:2220-2224 (1991); Wong-Staal *et al.,* PCT/US94/05700, and Rosenburg and Fauci (1993) in *Fundamental Immunology, Third Edition* Paul (ed) Raven Press, Ltd., New York

30   and the references therein, and Yu *et al., Gene Therapy* (1994) supra).

The present invention may also be used in the pharmaceutical industry. For example, it will provide information that eventually may enable cells from fetal tissue, which may the be transplanted into patients suffering from *e.g.* Parkinson's disease

5   or cancer, such as BCC. (For a brief review of methods of drug delivery, *see* Langer 249:1 527-1533 (1990), Remington's Pharmaceutical Sciences, Mack Publishing Company, Philadelphia, PA, 17th ed. (1985) *etc.*)

---

10   <u>Detailed description of the drawings</u>

Figure 1 shows SEQ ID NO 1, which is the amino acid sequence encoded by the novel human patched 2 gene.

Figure 2 shows SEQ ID NO 2, which is the nucleotide sequence encoding the pro-

15   tein disclosed by SEQ ID NO 1.

Figure 3 shows SEQ ID NO 3, wherein exons and introns are designated in the genomic sequence of the novel human patched 2 gene.

20   Figure 4A discloses an amino acid sequence comparison of the human PTCH2 (upper lines) and PTCH1 (lower lines) sequences. Vertical lines indicate identical amino acids, while dots similar amino acids. The PTCH2 sequence presented is composed of the original cDNA clones and of the products of the 5' RACE analysis.

25

Figure 4B is a representation of the alternative splicing events that result in different C-termini. In the parotid gland and the colon, the penultimate and the last exon are canonically joined together. In fetal brain however the penultimate exon with part of the 3' intron functions as the terminal exon. The intronic sequence is shown by

30   small letters with the flanking exonic by capital letters. Above the nucleotide sequence, the deduced amino acid sequence is shown, and below is the correspon-

ding sequence of the mouse Ptch2. The conserved intronic dinucleotides are shown by bold letters and the termination signals are indicated by asterisks. Note the absence of conservation of the position of the termination codons between the mouse and human PTCH2 sequences. The putative polyadenylation signals are also shown

5   in this diagram. The genomic organization was obtained by analyzing BAC clones encompassing the PTCH2 gene.

Figure 4C is a representation of the different variations of spliced transcripts encompassing exon 1 and exon 2 sequences. The canonical exons 1 and 2 are shown

10   by boxes and the intron between them by a solid line. The GT and AG dinucleotides spanning the sequences that are used as introns in individual transcripts are indicated by small letters. G, Genomic structure, derived from sequencing segements of BAC clones encompassing the PTCH2 gene; C, Canonical transcript; A, Transcript A (the skipped exons 9 and 10 of this product are not shown in the diagram); B,

15   Transcript B.

Figure 5A is a dark-field photomicrograph of a BCC tumor hybridized with ^{35}S-labeled antisense probe showing abundant signal for PTCH1 mRNA (light grains) in all BCC tumor cells.

20

Figure 5B discloses PTCH2 mRNA overexpression in BCC and is in contrast mainly expressed in the basaloid cells in the periphery of the tumor nests.

Figure 5C is another BCC showing a strong PTCH2 mRNA signal in the periphery

25   of the tumor nest (Tu), wheras no signal is detected in epidermis (Ep).

Figure 5D are sections of the same tumor (C) hybridized with the PTCH2 sense probe showed no signal.

30   Figure 5E shows immunoreactivity for Ki-67 (brown precipitate) seen in the periphery, in the cells that showed strong upregulation of PTCH2 mRNA.

Figure 5F discloses tumor nests under high power magnification demonstrate abundant PATCH2 mRNA signal (black grains) in the dark basaloid tumor cells and lower signal in the center (arrow). Bars (A-E), 24 μm, and F, 6 μm.

5

EXPERIMENTAL

Materials and methods

The RACE analysis was performed essentially as described before (Zaphiropoulos, P.G. and Toftgård, R. (1996): "cDNA cloning of a novel WD repeat protein map-

10   ping to the 9q22.3 chromosomal region", DNA Cell Biol. **15**, 1049-1056) using the Marathon kit (Promega). The primer sequences used for RACE are available upon request.

The PTCH2, 35S-labeled RNA probes used for the in situ hybridizations, that were

15   performed as previously described (Undén *et al.*, (1997), <u>supra</u>), corresponded to positions 218 to 437 and 838 to 920 in the PTCH2 sequence of Fig. 1A.

Results and discussion

In order to identify additional components of the PTCH/SHH cascade of signalling

20   events, the Incyte LifeSeq™ database (Incyte Pharmaceuticals Inc., Palo Alto, CA, USA) was searched using PTCH sequences. In addition to clones representing the PTCH cDNA, two nearly identical cDNAs were identified, from the parotid gland and the colon, that contained sequences similar to, but distinct from, the 3' end of PTCH. By 5' RACE analysis using fetal brain cDNAs additional sequence informa-

25   tion from these transcripts (termed PTCH2) and corresponding to a full length cDNA, was obtained (Fig. 4A). PTCH2 is 57% identical to PTCH1, with a significantly variable region present between the transmembrane domains 6 and 7, and 91% identical to the recently published mouse Ptch2 sequence (Motoyama, J., Takabatake, T., Takeshima, K. and Hui, C. (1998): "Ptch2, a second mouse Patched

30   gene is co-expressed with Sonic hedgehog", Nature Genet. **18**, 104-106). In simila-

rity with the mouse gene, PTCH2 lacks the C-terminal extension present in human, mouse and chicken PTCH1 (Goodrich, L.V., Johnson, R.L., Milenkovic, L., McMahon, J.A., and Scott, M.P. (1996): "Conservation of the hedgehog/patched signalling pathway from flies to mice: Induction of a mouse patched gene by Hedgehog", Genes Dev. **10**, 301-312, Marigo, V., Scott, M.P., Johnson, R.L., Goodrich, L.V. and Tabin, C.J. (1996): "Conservation in hedgehog signalling: Induction of a chicken patched homolog by Sonic hedgehog in the developing limb", Development **122**, 1225-1233). However, according to the present invention, it has been shown that the human PTCH2 cDNA terminates 36 amino acids earlier that the mouse Ptch2 sequence. Moreover, when 3' RACE was perfomed from fetal brain, an alternate C-terminal region was identified. This had a high structural similarity with the mouse Ptch2 C-terminal sequence and originates from the genomic region that links the last two exons of PTCH2 (Fig. 4B). Therefore, in these alternatively spliced transcripts, the penultimate exon with a segment of the contiguous 3' intron serves as the terminal exon.

Moreover the human and mouse transcripts differed in the position of the termination signals (the human sequence is 21 amino acids longer), suggesting a non-conserved, species-specific function of this alternate C-terminal domain. The finding of two possible C-terminal regions for PTCH2 is intriguing and implies a role of this phenomenon in modulating signalling. Additional alternatively spliced transcripts were also identified by the RACE analysis (Fig. 4C). Transcript A lacks the sequence that corresponds to exons 9 and 10 of PTCH1 (preliminary comparisons of the intron-exon junctions of PTCH2 with PTCH1 indicate a similar genomic organization), with the open reading frame being retained at the exon 8 to exon 11 junction. Exons 9 and 10 code for the last part of the first extracellular loop and for transmembrane domains 2 and 3 in the putative structure of the PTCH1 protein. Furthermore this transcript also lacks a 5' segment of the canonical exon 2, due to the use of an alternative 3' splice site present in this exon, with the open reading frame being maintained. The functional consequence of this alternative splicing is not yet known, but it is interesting to note that the extracellular loops in PTCH1 are

presumed to be involved in binding of the ligand SHH (Marigo *et al.*, (1996), Nature **384**, supra; Stone *et al.*, (1996), Nature **384**, supra) and that insertion of a neo-cassette in intron 9 of of the mouse PTCH1 gene is associated with a severe phenotype (Hahn, H., Wojnowski, L., Zimmer, A.M., Hall, J., Miller, G. and Zimmer, A.

5 (1998): "Rhabdomyosarcomas and radiation hypersensitivity in a mouse model of Gorlin syndrome", Nature Med. **4**, 619-622). Furthermore, exons 9 and 10 encode part of a putative sterol sensing domain (Osborne, T.F. and Rosenfeld, J.M. (1998): ~~"Related membrane domains in proteins of sterol sensing and cell signalling provide~~ a glimpse of treasures still buried within the dynamic realm of intracellular metabo-

10 lic regulation", Curr. Opin. Lipidol. **9**, 137-140, also found in PTCH1, and which has recently been implicated in mediating the potent modulating effect of cholesterol on SHH/PTCH signalling (Cooper, M.K., Porter, J.A., Young, K.E., and Beachy, P.A. (1998): "Teratogen-mediated inhibition of target tissue response to Shh signalling", Science **280**, 1603-1607). Thus, if PTCH2 also serves as a receptor for

15 SHH and/or related factors, the receptor form lacking exons 9 and 10 may show altered signalling properties. Transcript B contains additional sequences between canonical exons 1 and 2, that originate from the 5' end of intron 1. The open reading frame that includes the initiator methionine of exon 1 is not maintained in this transcript, suggesting that, if this transcript is functional, either the methionine in exon 2

20 or non-methionine codons are used in order to produce a protein product, in similarity to what has been proposed for the alternative spliced products of human PTCH1 (Hahn *et al.*, Cell **85**, supra). By radiation hybrid mapping the PTCH2 gene was localized to the short arm of chromosome 1, in difference to PTCH1 residing on chromosome 9q22.3.

25

The mouse and zebrafish homologs of PTCH2 have been reported to be expressed in a partly overlapping pattern with PTCH1 during embryonic development and to be induced by SHH (Motoyama *et al.*, (1998) Nature Genet. **18**, supra, Concordet, J.P., Lewis, K.E., Moore, J.W., Goodrich, L.V., Johnson, R.L., Scott, M.P., and

30 Ingham, P.W. (1996): "Spatial regulation of a zebrafish patched homologue reflects

the roles of sonic hedgehog and protein kinase A in a neural tube and somite patter-
ning", Development **122**, 2835-2846), implicating a role in this signalling pathway.
We were with this background interested to analyze the expression of PTCH2 in
BCCs which show consistent upregulation of PTCH1 in all tumor cells (Undén *et*
5   *al.*, (1997) Cancer res. **57**, <u>supra</u>). In situ hybridization was performed on six famili-
al and four sporadic BCCs of different histological subtypes. A strong positive sig-
nal for PTCH2 mRNA was observed exclusively in the tumor cells of all BCCs.
Notably, the signal was consistently stronger in the palisading peripheral cells of the
tumor nests (Fig. 4). These cells also showed a positive immunostaining for the cell
10   proliferation marker, Ki-67.

The finding that in BCCs having frequent mutations in the PTCH1 gene, the expres-
sion of the PTCH2 mRNAs is upregulated, tightly links the novel PTCH2 according
to the invention with the PTCH/SHH cascade of signalling events. It is therefore li-
15   kely that PTCH2 represents a target gene of this pathway which is under the negati-
ve regulation of PTCH1, precisely as PTCH1 itself. Moreover this observation
strongly suggests that PTCH2 has functions distinct from PTCH1 since upregulation
of PTCH2 expression appears unable to compensate for inactive PTCH1 protein.
This conclusion is also supported by the early embryonic lethality seen in PTCH1
20   (-/-) mice 5,13) and the lack of genetic heterogeneity in Gorlin syndrome. However,
whether PTCH2 may block the constitutive signalling of SMO, or could act as an
additional SHH receptor, possible dependent on alternative splicing, remains as the
subject of further experimentation.

25

CLAIMS

1. An isolated human protein or an analogue or variant thereof capable of participating in the human PTCH/SHH pathway (during embryonic development and/or carcinogenesis) comprising at least about 1040 amino acids of the sequence denoted SEQ ID NO. 1 of Figure 1.

2. A protein according to claim 1, which is essentially comprised of the sequence denoted denoted SEQ ID NO. 1 of Figure 1.

3. A nucleic acid encoding a protein according to any one of claims 1 and 2.

4. A nucleic acid encoding a protein according to claim 3 comprising at least about 3094 bases of the sequence denoted SEQ ID NO 2 of Figure 2.

5. An isolated genomic nucleic acid comprising parts or all of the sequence denoted SEQ ID NO 3 of Figure 3.

6. An isolated nucleic acid which comprises one o more mutations compared to the nucleic acid according to claim 5.

7. A nucleic acid having the sequence of any one of the splicing variants defined in Figure 4B.

8. An isolated nucleic acid capable of specifically hybridizing to a nucleic acid according to any one of claims 3-6.

9. A protein or polypeptide encoded by a nucleic acid according to claim 7 or 8.

10. A vector comprising a nucleic acid according to any one of claim 3-8.

11. A recombinant cell comprising a vector according to claim 10.

12. An antibody which specifically binds to a protein according to claim 1, 2 or 9.

5

13. A recombinant cell expressing an antibody according to claim 12.

14. A kit for the detection of a human PTCH2 gene or polypeptide comprising in a container a molecule selected from the group consisting of a nucleic acid according to any one of claims 3-8, a polypeptide or protein according to claim 1, 2 or 9 or an antibody according to claim 12.

10

15. Use of a nucleic acid selected from the group consisting of SEQ ID NO. 2 and SEQ ID NO. 3 in gene therapy.

15

ABSTRACT

The present invention relates to a novel human patched-like gene (PTCH2), which
for the first time has been cloned and sequenced. Several alternatively spliced

5   mRNA forms of PTCH2 have been ideintified, including transcripts lacking seg-
ments thought to be involved in sonic hedgehog (SHH) binding and mRNAs with
differentially defined 3' terminal exons. Further, the invention also relates to the
protein encoded by the present PTCH2 as well as to functional analogues and vari-
ants thereof.

10

Patched 2 amino acid sequence   SEQ ID NO 1

```
   1  MTRSPPLREL  PPSYTPPART  AAPQILAGSL  KAPLWLRAYF  QGLLFSLGCG

  51  IQRHCGKVLF  LGLLAFGALA  LGLRMAIIET  NLEQLWVEVG  SRVSQELHYT

 101  KEKLGEEAAY  TSQMLIQTAR  QEGENILTPE  ALGLHLQAAL  TASKVQVSLY

 151  GKSWDLNKIC  YKSGVPLIEN  GMIERMIEKL  FPCVILTPLD  CFWEGAKLQG

 201  GSAYLPGRPD  IQWTNLDPEQ  LLEELGPFAS  LEGFRELLDK  AQVGQAYVGR

 251  PCLHPDDLHC  PPSAPNHHSR  QAPNVAHELS  GGCHGFSHKF  MHWQEELLLG

 301  GMARDPQGEL  LRAEALQSTF  LLMSPRQLYE  HFRGDYQTHD  IGWSEEQAST

 351  VLQAWQRRFV  QLAQEALPEN  ASQQIHAFSS  TTLDDILHAF  SEVSAARVVG

 401  GYLLMLAYAC  VTMLRWDCAQ  SQGSVGLAGV  LLVALAVASG  LGLCALLGIT

 451  FNAATTQVLP  FLALGIGVDD  VFLLAHAFTE  ALPGTPLQER  MGECLQRTGT

 501  SVVLTSINNM  AAFLMAALVP  IPALRAFSLQ  AAIVVGCTFV  AVMLVFPAIL

 551  SLDLRRRHCQ  RLDVLCCFSS  PCSAQVIQIL  PQELGDGTVP  VGIAHLTATV

 601  QAFTHCEASS  QHVVTILPPQ  AHLVPPPSDP  LGSELFSPGG  STRDLLGQEE

 651  ETRQKAACKS  LPCARWNLAH  FARYQFAPLL  LQSHAKAIVL  VLFGALLGLS

 701  LYGATLVQDG  LALTDVVPRG  TKEHAFLSAQ  LRYFSLYEVA  LVTQGGFDYA

 751  HSQRALFDLH  QRFSSLKAVL  PPPATQAPRT  WLHYYRNWLQ  GIQAAFDQDW

 801  ASGRITRHSY  RNGSEDGALA  YKLLIQTGDA  QELLDFSQLT  TRKLVDREGL

 851  IPPELFYMGL  TVWVSSDPLG  LAASQANFYP  PPPEWLHDKY  DTTGENFRIP

 901  PAQPLEFAQF  PFLLRGLQKT  ADFVEAIEGA  RAACAEAGQA  GVHAYPSGSP

 951  FLFWEQYLGL  RRCFLLAVCI  LLVCTFLVCA  LLLLNPWTAG  LIVLVLAMMT

1001  VELFGIMGFL  GIKLSAIPVV  ILVASVGIGV  EFTVHVALGF  LTTQGSRNLR

1051  AAHALEHTFA  PVTDGAISTL  LGLLMLAGSH  FDFIVRYFFA  ALTVLTLLGL

1101  LHGLVLLPVL  LSILGPPPEV  IQMYKESPEI  LSPPAPQGGG  LRPEEI
```

Fig. 1

Patched 2 nucleotide sequence   SEQ ID NO 2

```
   1  ATGACTCGAT CGCCGCCCCT CAGAGAGCTG CCCCCGAGTT ACACACCCCC

  51  AGCTCGAACC GCAGCACCCC AGATCCTAGC TGGGAGCCTG AAGGCTCCAC

 101  TCTGGCTTCG TGCTTACTTC CAGGGCCTGC TCTTCTCTCT GGGATGCGGG

 151  ATCCAGAGAC ATTGTGGCAA AGTGCTCTTT CTGGGACTGT TGGCCTTTGG

 201  GGCCCTGGCA TTAGGTCTCC GCATGGCCAT TATTGAGACA AACTTGGAAC

 251  AGCTCTGGGT AGAAGTGGGC AGCCGGGTGA GCCAGGAGCT GCATTACACC

 301  AAGGAGAAGC TGGGGGAGGA GGCTGCATAC ACCTCTCAGA TGCTGATACA

 351  GACCGCACGC CAGGAGGGAG AGAACATCCT CACACCCGAA GCACTTGGCC

 401  TCCACCTCCA GGCAGCCCTC ACTGCCAGTA AAGTCCAAGT ATCACTCTAT

 451  GGGAAGTCCT GGGATTTGAA CAAAATCTGC TACAAGTCAG GAGTTCCCCT

 501  TATTGAAAAT GGAATGATTG AGCGGATGAT TGAGAAGCTG TTTCCGTGCG

 551  TGATCCTCAC CCCCCTCGAC TGCTTCTGGG AGGGAGCCAA ACTCCAAGGG

 601  GGCTCCGCCT ACCTGCCCGG CCGCCCGGAT ATCCAGTGGA CCAACCTGGA

 651  TCCAGAGCAG CTGCTGGAGG AGCTGGGTCC CTTTGCCTCC CTTGAGGGCT

 701  TCCGGGAGCT GCTAGACAAG GCACAGGTGG CCAGGCCTA CGTGGGGCGG

 751  CCCTGTCTGC ACCCTGATGA CCTCCACTGC CCACCTAGTG CCCCCAACCA

 801  TCACAGCAGG CAGGCTCCCA ATGTGGCTCA CGAGCTGAGT GGGGGCTGCC

 851  ATGGCTTCTC CCACAAATTC ATGCACTGGC AGGAGGAATT GCTGCTGGGA

 901  GGCATGGCCA GAGACCCCCA AGGAGAGCTG CTGAGGGCAG AGGCCCTGCA

 951  GAGCACCTTC TTGCTGATGA GTCCCCGCCA GCTGTACGAG CATTTCCGGG

1001  GTGACTATCA GACACATGAC ATTGGCTGGA GTGAGGAGCA GGCCAGCACA

1051  GTGCTACAAG CCTGGCAGCG GCGCTTTGTG CAGCTGGCCC AGGAGGCCCT

1101  GCCTGAGAAC GCTTCCCAGC AGATCCATGC CTTCTCCTCC ACCACCCTGG

1151  ATGACATCCT GCATGCGTTC TCTGAAGTCA GTGCTGCCCG TGTGGTGGGA

1201  GGCTATCTGC TCATGCTGGC CTATGCCTGT GTGACCATGC TGCGGTGGGA
```

Fig. 2

```
1251   CTGCGCCCAG TCCCAGGGTT CCGTGGGCCT TGCCGGGGTA CTGCTGGTGG

1301   CCCTGGCGGT GGCCTCAGGC CTTGGGCTCT GTGCCCTGCT CGGCATCACC

1351   TTCAATGCTG CCACTACCCA GGTGCTGCCC TTCTTGGCTC TGGGAATCGG

1401   CGTGGATGAC GTATTCCTGC TGGCGCATGC CTTCACAGAG GCTCTGCCTG

1451   GCACCCCTCT CCAGGAGCGC ATGGGCGAGT GTCTGCAGCG CACGGGCACC

1501   AGTGTCGTAC TCACATCCAT CAACAACATG GCCGCCTTCC TCATGGCTGC

1551   CCTCGTTCCC ATCCCTGCGC TGCGAGCCTT CTCCCTACAG GCGGCCATAG

1601   TGGTTGGCTG CACCTTTGTA GCCGTGATGC TTGTCTTCCC AGCCATCCTC
```

```
1651   AGCCTGGACC TACGGCGGCG CCACTGCCAG CGCCTTGATG TGCTCTGCTG

1701   CTTCTCCAGT CCCTGCTCTG CTCAGGTGAT TCAGATCCTG CCCCAGGAGC

1751   TGGGGGACGG GACAGTACCA GTGGGCATTG CCCACCTCAC TGCCACAGTT

1801   CAAGCCTTTA CCCACTGTGA AGCCAGCAGC CAGCATGTGG TCACCATCCT

1851   GCCTCCCCAA GCCCACCTGG TGCCCCCACC TTCTGACCCA CTGGGCTCTG

1901   AGCTCTTCAG CCCTGGAGGG TCCACACGGG ACCTTCTAGG CCAGGAGGAG

1951   GAGACAAGGC AGAAGGCAGC CTGCAAGTCC CTGCCCTGTG CCCGCTGGAA

2001   TCTTGCCCAT TTCGCCCGCT ATCAGTTTGC CCCGTTGCTG CTCCAGTCAC

2051   ATGCTAAGGC CATCGTGCTG GTGCTCTTTG GTGCTCTTCT GGGCCTGAGC

2101   CTCTACGGAG CCACCTTGGT GCAAGACGGC CTGGCCCTGA CGGATGTGGT

2151   GCCTCGGGGC ACCAAGGAGC ATGCCTTCCT GAGCGCCCAG CTCAGGTACT

2201   TCTCCCTGTA CGAGGTGGCC CTGGTGACCC AGGGTGGCTT TGACTACGCC

2251   CACTCCCAAC GCGCCCTCTT TGATCTGCAC CAGCGCTTCA GTTCCCTCAA

2301   GGCGGTGCTG CCCCCACCGG CCACCCAGGC ACCCCGCACC TGGCTGCACT

2351   ATTACCGCAA CTGGCTACAG GGAATCCAGG CTGCCTTTGA CCAGGACTGG

2401   GCTTCTGGGC GCATCACCCG CCACTCGTAC CGCAATGGCT CTGAGGATGG

2451   GGCCCTGGCC TACAAGCTGC TCATCCAGAC TGGAGACGCC CAGGAGCTTC

2501   TGGATTTCAG CCAGCTGACC ACAAGGAAGC TGGTGGACAG AGAGGGACTG
```

```
2551   ATTCCACCCG AGCTCTTCTA CATGGGGCTG ACCGTGTGGG TGAGCAGTGA
```

Fig. 2 (forts)

```
2601   CCCCCTGGGT CTGGCAGCCT CACAGGCCAA CTTCTACCCC CCACCTCCTG

2651   AATGGCTGCA CGACAAATAC GACACCACGG GGGAGAACTT TCGCATCCCG

2701   CCAGCTCAGC CCTTGGAGTT TGCCCAGTTC CCCTTCCTGC TGCGTGGCCT

2751   CCAGAAGACT GCAGACTTTG TGGAGGCCAT CGAGGGGGCC CGGGCAGCAT

2801   GCGCAGAGGC CGGCCAGGCT GGGGTGCACG CCTACCCCAG CGGCTCCCCC

2851   TTCCTCTTCT GGGAACAGTA TCTGGGCCTG CGGCGCTGCT TCCTGCTGGC

2901   CGTCTGCATC CTGCTGGTGT GCACTTTCCT CGTCTGTGCT CTGCTGCTCC

2951   TCAACCCCTG GACGGCTGGC CTCATAGTGC TGGTCCTGGC GATGATGACA

3001   GTGGAACTCT TTGGTATCAT GGGTTTCCTG GGCATCAAGC TGAGTGCCAT

3051   CCCCGTGGTG ATCCTTGTGG CCTCTGTAGG CATTGGCGTT GAGTTCACAG

3101   TCCACGTGGC TCTGGGCTTC CTGACCACCC AGGGCAGCCG GAACCTGCGG

3151   GCCGCCCATG CCCTTGAGCA CACATTTGCC CCCGTGACCG ATGGGGCCAT

3201   CTCCACATTG CTGGGTCTGC TCATGCTTGC TGGTTCCCAC TTTGACTTCA

3251   TTGTAAGGTA CTTCTTTGCG GCGCTGACAG TGCTCACGCT CCTGGGCCTC

3301   CTCCATGGAC TCGTGCTGCT GCCTGTGCTG CTGTCCATCC TGGGCCCGCC

3351   GCCAGAGGTG ATACAGATGT ACAAGGAAAG CCCAGAGATC CTGAGTCCAC

3401   CAGCTCCACA GGGAGGCGGG CTTAGGCCCG AGGAGATC**TA G**
```

Fig. 2 (forts)

Patched 2 Intron-Exon organization   SEQ ID NO 3

The intron sequences between exons 2 - 3 and exons 18 - 19 are
missing (introns: small letters, exons: capital letters). Small
letters in the first exon indicate nucleotides that have not been
unambigouisly determined.
The start and termination codons are indicated.

Exon 1

```
   1   CGGGTGAATC CCGGCGCCGC GCCCCGGACC CGCAGCTCCC TGCACTCCTC

  51   CCTCCCAGCC GCTTTAACAC CCACACCCCA CAGTCTCTCC CACGsCCGCG

 101   CCTTGGCGGC CCCACTGAAT CCCTACGCGG GGCCCAGCGG TACCGGGAGA

 151   CCGGGCTAGC CTATGGGAGC GCCCAGATAA CGCGGGTTGG GGGCGCCCGC

 201   GCCCcCATCC CCGCCAGCAT GACTCGATCG CCGCCCCTCA GAGAGCTGCC

 251   CCCGAGTTAC ACACCCCAG CTCGAACCGC AGCACCCCAG gtgagtagag

 301   ggggagctgg aagaaggaag agagcggagc caggtctgtc actcgggcct

 351   ctgcaaggtt tgtgatgtct tgaagtgccg agtgtcatta gatgtctgaa

 401   ggcaagtgag agccagcacc gcaagcaagt tgtgcgtgtg tgtcggtgtg

 451   tctgtgccgg tgtctcctca tcgtctggcc agtgagaatg aatgtctgtg

 501   ggttcacctc tgtgtccacc cgacgacagg tgtgtgtaca tatgtatcct

 551   gctctcagaa aatgggccta tgccgccggg cgcggtgact cacgcctgta

 601   atcccaacac tgggaggctg aggcaggcag attacctgag gtcaggagtt⁻

 651   cgagaccagc caggccaaca tggggaaact ctgtctctac taaaaataaa

 701   aattagcagg gcgtggtggc gggcgcctgt agtcccaact actcgggagg

 751   ctgaggcagg agaatctctt gaacctggga ggcggaggtt gcagtcaagc

 801   cgagatcaca ccactgcact ccagccaggg caacagagcg agatgcgtct

 851   caaaaaaaaa aaaaaaaaaa aaaaggagag aaaacaaaaa gaaaagaaag

 901   gaaaataggc ctatgccttc ctcaggtgtg tgctggggat ggtgggtgtt

 951   acatcttcca agtctgggcc tgtgtctgtg ttggtgctcc ctgtcccaca

1001   tccagaaatc aagaagcgag ggctgggcag cagatataca gggtgagaag
```

Fig. 3

1051 ggaaggattt catgcattgt tacagtgatg cctggctgac ccttctcttt
           EXON 2
1101 ccatcccagA TCCTAGCTGG GAGCCTGAAG GCTCCACTCT GGCTTCGTGC

1151 TTACTTCCAG GGCCTGCTCT TCTCTCTGGG ATGCGGGATC CAGAGACATT

1201 GTGGCAAAGT GCTCTTTCTG GGACTGTTGG CCTTTGGGGC CCTGGCATTA

1251 GGTCTCCGCA TGGCCATTAT TGAGACAAAC TTGGAACAGC TCTGGGTAGA

1301 AGTGGGCAGC CGGGTGAGCC AGGAGCTGCA TTACACCAAG GAGAAGCTGG

1351 GGGAGGAGGC TGCATACACC TCTCAGATGC TGATACAGAC CGCACGCCAG

1401 GAGGGAGAGA ACATCCTCAC ACCCGAAGCA CTTGGCCTCC ACCTCCAGGC

1451 AGCCCTCACT GCCAGTAAAG TCCAAGTATC ACTCTATGGG AAG.......

1501 ....g

1551 tgagtctggc tgagcccctg agcagctggg ggcgaggcgt gctgtggggg

1601 ttctggagtg ggaatcccct tcttctgctg atctcctatg cccctggcta
           EXON 4
1651 ttgcagTCCT GGGATTTGAA CAAAATCTGC TACAAGTCAG GAGTTCCCCT

1701 TATTGAAAAT GGAATGATTG AGCGGgtaag tgtcctgaga gggagtagag

1751 gcagaacttt ttctgtagcg tgggaggact cagagaccga gcaagcccca

1801 cagcctgcaa tctgccccct taaaactaag gaggggggatt gcagagggca

1851 tcctacaaag gttgtggggc aggactgacg tggcccgggg tatccctggc
           EXON 5
1901 agATGATTGA GAAGCTGTTT CCGTGCGTGA TCCTCACCCC CCTCGACTGC

1951 TTCTGGGAGG GAGCCAAACT CCAAGGGGGC TCCGCCTACC TGCCgtgagt

2001 gccactcctg gggccctgct tcatctcccg ctggggactc tcccagcaga

2051 aaggaggggt ctggggaatg aggatgatca aaaccttacc aaggtcctaa

2101 ttacctccca ggccaggaac agagagcatg ggcttcccca aggctctctc

2151 cacatcctcc ttctctttcc ctctcaagga aggaagacct gacttattta

2201 cacaaaacta aacacaaaga tctgtaagat ctgagcaaag gagaaaaaga

2251 tccccacaaa gaggctttgc tggggggaaat tacctaggtg tttgctaagc

2301 cattgcccag gccagaaaga aaacctgcta caggcatgtg cctgctggtt

2351 gtatattaga accaagcaca cagcttggta aggaactcag tggggccttt

Fig 3 (forts)

```
2401 ctgggcccct tctatgtatt aggtaacccct gccctgatat tcgtctcagc

2451 cccttgtact cttctacagc tcactgtagc accctggtgg gcccatgcag

2501 cctggcagtt ctgagaagct gaggcttgca caccctccat atggaaggac

2551 aaatcggcag ataagaggag ggtggggtac agcatggcgc cccagcagca

2601 gtttggagcc tgggttttcg tccctgaccc tcaccaacta taggcttttc
                 EXON 6
2651 cctcagCGGC CGCCCGGATA TCCAGTGGAC CAACCTGGAT CCAGAGCAGC

2701 TGCTGGAGGA GCTGGGTCCC TTTGCCTCCC TTGAGGGCTT CCGGGAGCTG

2751 CTAGACAAGG CACAGGTGGG CCAGGCCTAC GTGGGGCGGC CCTGTCTGCA
```

```
2801 CCCTGATGAC CTCCACTGCC CACCTAGTGC CCCCAACCAT CACAGCAGGC

2851 AGgtgggttc caaccaggtc tgccagggaa aggctgtttt ccttcccttt

2901 cccttcctca tactcctgtg ttctggggga gctgactgct ctgtgccctg

2951 accccccact tcctggccat tattaccctg ctcccacagt gccaggcccc

3001 caatgttcca ttcccattca gttatcctac ggagccctca agtggtatat

3051 atgaatccct ttttcctttt ctaagcctag ataaggctgg acttcttttt

3101 tttttttttt ttgagtctca ctctgtcacc caggctggag tgcagtagtt

3151 cgatcttggc tcactgcaac ctcggctcaa gcaattctcc tgccttagcc

3201 tcctgagtag ctgggattac aggtgcccac caccatgccc ggctaatttt

3251 tattagcctc ccaaagtgct gggattacag gcgtgagcca ctgcgcctgg

3301 ccaaggctgg acttttttatc aaaatagact aatacaggga aactaagaac

3351 acagcaggta agcatgaata tcatacctgg tttcccaggt ttctttgtgg

3401 ccctgcaaat gtggtacttt tttcagaatc cgccagttac accagctcct

3451 cccagaagcc tacttccagg cctctgcttc cccttggggc ttcctgtctg

3501 cgggatacta gctgttcact cctgcagagc agtcaagagg ctcagaatag

3551 ttacctacac tccagcccta ctgagcttca tggcagcgtg gttcctggag

3601 gtggaagccc agggacactc agttatccac ggccagggcc ttgagcatta
                                      EXON 7
3651 acccctcctg ttcccctcca gGGCTCCCAA TGTGGCTCAC GAGCTGAGTG
```

```
3701 GGGGCTGCCA TGGCTTCTCC CACAAATTCA TGCACTGGCA GGAGGAATTG
```

Fig 3 (forts)

```
3751  CTGCTGGGAG GCATGGCCAG AGACCCCCAA GGAGAGCTGC TGAGgtaggg

3801  tctcctctgg gagttggtga ggggactctg ttcatgagaa cccatactgt

3851  aatgccaggc agctctggca aaaggccctt cacatccctc accaggtgtt
                                                        EXON 8
3901  tgggccagct ctgacccctg gttctcccac acccccacca gGGCAGAGGC

3951  CCTGCAGAGC ACCTTCTTGC TGATGAGTCC CCGCCAGCTG TACGAGCATT

4001  TCCGGGGTGA CTATCAGACA CATGACATTG GCTGGAGTGA GGAGCAGGCC

4051  AGCACAGTGC TACAAGCCTG GCAGCGGCGC TTTGTGCAGg tcggtatgga

4101  caaggacaag gggggtgccc tgaggccatt ccctcctcct gcccccctcct
                                               EXON 9
```

```
4151  atccaccctg tttctccagC TGGCCCAGGA GGCCCTGCCT GAGAACGCTT

4201  CCCAGCAGAT CCATGCCTTC TCCTCCACCA CCCTGGATGA CATCCTGCAT

4251  GCGTTCTCTG AAGTCAGTGC TGCCCGTGTG GTGGGAGGCT ATCTGCTCAT

4301  Ggtgggtctt gcacctggca ccttgccccc accccacctc caaccagtgc
                                                        EXON 10
4351  ccaccctggg agcccctgag actgcccttt ccccccacag CTGGCCTATG

4401  CCTGTGTGAC CATGCTGCGG TGGGACTGCG CCCAGTCCCA GGGTTCCGTG

4451  GGCCTTGCCG GGGTACTGCT GGTGGCCCTG GCGGTGGCCT CAGGCCTTGG

4501  GCTCTGTGCC CTGCTCGGCA TCACCTTCAA TGCTGCCACT ACCCAGgtac

4551  gccaggactg cagggcagac tcagtgccag tcaccaggct tcacgggtcc
                                          EXON 11
4601  tcagctgccc gctcctctgc ccctccagGT GCTGCCCTTC TTGGCTCTGG_

4651  GAATCGGCGT GGATGACGTA TTCCTGCTGG CGCATGCCTT CACAGAGGCT

4701  CTGCCTGGCA CCCCTCTCCA Ggtggggcct tgtcccccag ggctcatctg

4751  aggcagctca gcttactggt taagagcctc ttggttcaag tgacccttgg

4801  gctgctaatg aacctcggtg cctcttgtcc ccatctgtaa acaggggaaa

4851  taatagtgct gtgtcctaag ggttattgtt tggatcagtg aggtaactca

4901  agttgaatgc ttagaacagc ccatcatacg tacatggtac ccaataaatg

4951  ctagccactg tgttatgact gccccacctc tgcaccccaa gttcctgagc

5001  ctccccttca ctccactttg acacggcccc tcccttgtga cctgagggca
                                    EXON 12
```

```
5051  ggtccccact ctgtcctggc agGAGCGCAT GGGCGAGTGT CTGCAGCGCA
```

Fig 3 (forts)

```
5101  CGGGCACCAG TGTCGTACTC ACATCCATCA ACAACATGGC CGCCTTCCTC

5151  ATGGCTGCCC TCGTTCCCAT CCCTGCGCTG CGAGCCTTCT CCCTACAGGC

5201  GGCCATAGTG GTTGGCTGCA CCTTTGTAGC CGTGATGCTT GTCTTCCCAG

5251  CCATCCTCAG CCTGGACCTA CGGCGGCGCC ACTGCCAGCG CCTTGATGTG

5301  CTCTGCTGCT TCTCCAGgta ctgcgtgcgc cccagccccct tcctcccgtg

5351  acccacgcca gcctgtcccc tcaccagcat ttcaaggcac agacctgtca
                                          EXON 13
5401  tccactctct acctcttcca gTCCCTGCTC TGCTCAGGTG ATTCAGATCC

5451  TGCCCCAGGA GCTGGGGGAC GGGACAGTAC CAGTGGGCAT TGCCCACCTC

5501  ACTGCCACAG TTCAAGCCTT TACCCACTGT GAAGCCAGCA GCCAGCATGT

5551  GGTCACCATC CTGCCTCCCC AAGCCCACCT GGTGCCCCCA CCTTCTGACC

5601  CACTGGGCTC TGAGCTCTTC AGCCCTGGAG GGTCCACACG GGACCTTCTA

5651  GGCCAGGAGG AGGAGACAAG GCAGAAGGCA GCCTGCAAGT CCCTGCCCTG

5701  TGCCCGCTGG AATCTTGCCC ATTTCGCCCG CTATCAGTTT GCCCCGTTGC

5751  TGCTCCAGTC ACATGCTAAG gtaagactgg gcagagcagg gcagagactt

5801  agcatctctg ggcccagaag ggcagagagg gcttagtcca ctgcctgagg
                                                          EX
5851  ggctgggggc agccctgggg tctccagctt agttgctaca tcccgcagGC
        XON 14
5901  CATCGTGCTG GTGCTCTTTG GTGCTCTTCT GGGCCTGAGC CTCTACGGAG

5951  CCACCTTGGT GCAAGACGGC CTGGCCCTGA CGGATGTGGT GCCTCGGGGC_

6001  ACCAAGGAGC ATGCCTTCCT GAGCGCCCAG CTCAGGTACT TCTCCCTGTA

6051  CGAGGTGGCC CTGGTGACCC AGGGTGGCTT TGACTACGCC CACTCCCAAC

6101  GCGCCCTCTT TGATCTGCAC CAGCGCTTCA GTTCCCTCAA GGCGGTGCTG

6151  CCCCCACCGG CCACCCAGGC ACCCCGCACC TGGCTGCACT ATTACCGCAA

6201  CTGGCTACAG Ggtgagaggc gaggagacgg gcagggaggg gtgctgcagg

6251  gagaaacgcc ctggggccac cagctaatag aaccctatcc tggtctcccc
         EXON 15
6301  cagGAATCCA GGCTGCCTTT GACCAGGACT GGGCTTCTGG GCGCATCACC

6351  CGCCACTCGA CCGCAATGGC TCTGAGGATG GGGCCCTGGC CTACAAGCTG

6401  CTCATCCAGA CTGGAGACGC CCAGGAGCTT CTGGATTTCA GCCAGgttgg
```

Fig 3 (forts)

```
6451   gagagggctg gaggggtcca ctagtacagg ggctgcaggc ctcctgggcc
                                                    EXON 16
6501   caggccttca gccctctctg cctctgcagC TGACCACAAG GAAGCTGGTG
6551   GACAGAGAGG GACTGATTCC ACCCGAGCTC TTCTACATGG GGCTGACCGT
6601   GTGGGTGAGC AGTGACCCCC TGGGTCTGGC AGCCTCACAG GCCAACTTCT
6651   ACCCCCCACC TCCTGAATGG CTGCACGACA AATACGACAC CACGGGGGAG
6701   AACTTTCGCA gtgagtcttg gggggagctc ggcaagagcc tcagcctcgc
6751   ccacacaagc cctgagcctg aggccctgcc cactctgccc cgtgctcacc
                                                    EXON 17
6801   gccctgtccc tctccctctt ctcccttccc ctcccctcca cagTCCCGCC
6851   AGCTCAGCCC TTGGAGTTTG CCCAGTTCCC TTTCCTGCTG CGTGGCCTCC
6901   AGAAGACTGC AGACTTTGTG GAGGCCATCG AGGGGGCCCG GGCAGCATGC
6951   GCAGAGGCCG GCCAGGCTGG GGTGCACGCC TACCCCAGCG GCTCCCCCTT
7001   CCTCTTCTGG GAACAGTATC TGGGCCTGCG GCGCTGCTTC CTGCTGGCCG
7051   TCTGCATCCT GCTGGTGTGC ACTTTCCTCG TCTGTGCTCT GCTGCTCCTC
7101   AACCCCTGGA CGGCTGGCCT CATAgtgagt gcttgcagga gtggggacag
7151   agacacccca cccttccctg cccagcctgt catccctcct gccaggagcc
                                                    EXON 18
7201   ctctgtgagc cctgtctccc tcagGTGCTG GTCCTGGCGA TGATGACAGT
7251   GGAACTCTTT GGTATCATGG GTTTCCTGGG CATCAAGCTG AGTGCCATCC
7301   CCGTGGTGAT CCTTGTGGCC TCTGTAGGCA TTGGCGTTGA GTTCACAGTC_
7351   CACGTGGCTC TGGGCTTCCT GACCACCCAG GGCAGCCGGA ACCTGCGGGC
7401   CGCCCATGCC CTTGAGCACA CATTTGCCCC CGTGACCGAT GGGGCCATCT
7451   CCACATTGCT GGGTCTGCTC ATGCTTGCTG GTTCCCACTT TGACTTCATT
7501   GTAAG..... .......... .......... .......... ..........
7551   .......... gtagggaggg ctcggggcag ggaggcaggg ctcaggacag
                                                    EXON 20
7601   gcctgggctg actcccccca caccctaccc ctagGTACTT CTTTGCGGCG
7651   CTGACAGTGC TCACGCTCCT GGGCCTCCTC CATGGACTCG TGCTGCTGCC
7701   TGTGCTGCTG TCCATCCTGG GCCCGCCGCC AGAGgtgacc acaccctcgg
7751   caccatccct ctactcccag cccaagggac ggggtaggga gaggcaaggg
```

Fig 3 (forts)

```
7801   aagggacaga   gccctgtggc   ccacagacag   gtacctcccc   aacaggtgcc

7851   accagctgaa   ggtggcagcc   tcctcctttc   cccagacacc   atgttcctgc

7901   ccctcagccc   tcctggcttc   ttcatgggac   ccaccttaga   cttttaggat

7951   ccagaacaag   gtgcagggtt   tgccccaggc   ctcaacatcc   tgtcgcctgc

8001   cagctctcat   atcctgctgg   agaccaacaa   gggccccagc   ttcccaacag

8051   tcatggtaat   ccccagcgag   atgctaaagg   ggacgggagc   cccagggggcc
                                                        EXON 21
8101   cgtgggctta   ctgggctgg   tgtctcccca   cagGTGATAC   AGATGTACAA

8151   GGAAAGCCCA   GAGATCCTGA   GTCCACCAGC   TCCACAGGGA   GGCGGGCTTA
```
```
8201   Ggtgggggc   atcctcctcc   ctgccccaga   gctttgccag   agtgactacc

8251   tccatgaccg   tggccatcca   cccaccccc   ctgcctggtg   cctacatcca

8301   tccagcccct   gatgagcccc   cttggtcccc   tgctgtcact   agctctggca

8351   acctcagttc   caggggacca   ggtccagcca   ctgggtgaaa   gagcagctga

8401   agcacagaga   ccatgtgtgg   ggcgtgtggg   gtcactggga   agcactgggt

8451   ctggtgttag   acgcaggatg   gacccctgga   gggctctgct   gctgctgcat

8501   cccctctccc   gacccagctg   tcatgggcct   ccctgatatc   catacagaac

8551   agccaccgat   ttgcacatcc   aggcctgtgt   gagcctgtat   ctgtgtcact

8601   tgagagtgaa   agctggcact   tggggctgca   gtgcagccct   gtcccccttc

8651   ccaccccaca   ccactgcctg   cccagctgac   caagcctgag   ggaccctcca

8701   gcacccttcc   gtctggtgac   tcctgggcag   gctctccata   tccctgccca

8751   cctcctacca   catccattat   ttatatgaaa   atgtctattt   ttgtagtata

8801   catacatgtt   agctatgatg   aaagttttat   tttttaaaga   atgaaatata

8851   ttctatgtga   agctatgatg   aaagttttat   tttttaaaga   atgaaatata

8901   ttctatgtga   actaatctcg   aaagttttat   tttttaaaga   atgaaatata

8951   ttctatgtgt   gcaagtgaac   attagcttca   gttgcttttt   tttggacaga

9001   gtggggagtt   tgcaagtgaa   cattagctat   tggaaggagc   ttctctggtg

9051   ccaggacctg   aggtattagc   ttctctagtt   ctgggtggaa   aagaccccag
```
```
9101   attctggatt   tttgtcatat   acttggtaac   atcatctgga   ttaagtgctt
```

Fig 3 (forts)

```
9151   actatacaaa acgataacaa attttgttgg tgtgaaatcc tactgggttc

9201   aatctggaga ccgagagcag aaaaaaaaga accccactgt gtggctttca

9251   gagccaccat attccagcct gcccgtctct ccagactcac ctccacctac

9301   ctgcttcacc cgcacgggaa acggcaaggc agaggggcaa agccatgcag

9351   caggtggaag gcgaggtgga ggcagatcag gaaagcagcc agttgaagca

9401   gagagaggtc aacagggtct ggggagcttc tcaggaggtt tgtggaccca

9451   gggaaaggag ccaggttcca gagcaacctc caaggcaaag gcctctgtaa

9501   gttggttgtc ctgacagccg agaggtgtct ttggccagtc agccagtgga

9551   tcagttgcgg gaactgctca gaaactgagg tgctagcagt tagtgaggac

9601   acagcgtaag ttgtttgttc tgtgaaagtt gaacagctcc actaagcaga

9651   ggccttgaag agtggccaca gccctggaat agagcacaga gcctcaccta

9701   gaggcgtggg gaggtttgca actgcccctt cccagccata gcttaggacc

9751   catagtctag ttcacataga ccctgggctc caaccaccca ctcaccagga

9801   atgatcccac cccaggaaca atgcgttctc acatcccacc ccacctggac

9851   aaaggccagg aaatcatgtt ctgaccaaaa gatacaacaa caaaaacaac

9901   aacaacaaaa aacgcctatt gcaattgaat ccacgctaaa atgcctaaaa

9951   agctcaagag aagcgggtag ttggcagaga acctagagta gggggtgcaa

10001  ccagcaggcc caagggaggg aggctgcatt tgggtccagc agtgtttggg

10051  tcaccaagaa gggccttcta ggtggagcag agagagctca ccaggccaga

10101  atagtgcaaa gggggtcagc cctcagtgcc acttaccagc ggagtaaccc
                                                          E
10151  tgggcaagtt agccagcctc actaagcctc cccatcttca tctttccagG
       XON 22
10201  CCCGAGGAGA TCTAGCCTCT GCCTCCCACC CCAGCACCCC CTCATCAGAC

10251  ACAAGGAGCG CCACTGTCTG GACAGGCTGA ATTGGTCTTC GGGTCCCTAA

10301  TTTCTCATAC GCCATTCCCT CTGCCTAGAA CACTTTCTCA CCTCCCCTTG

10351  ATGTGACCCC ATATCACCCT TCGAGGTGAA TTGGATCGGA TGCCATCTCC

10401  TCCAGGAGGG GTGGGGTCGT GCCTCCTGTG AGGTCCCAGT GCCCCTGAGT

10451  GTCTGTGCCC GTCTGTTTCC CCGTCCCTCT CTCTAAGCCC GGAGGCTTAC
```

Fig 3 (forts)

10501 TGCGGGTAAG GACGGCGGGA CAGGACCTTA ACCCCTGGGA CGAACACCAG

10551 CTCCGCAAAG GACTCCGCAC CCGGCGCCGC CCACGGGGTG CGGGTCCCAG

10601 GAGGACCAGC AGAGAGGAGC ATAGGAGAGC AAAGGAGATC AGTGACCCAT

10651 GGCTTCCCCG GTGGCGCGGA ACAGCCCGGA GCCGCCTGTG ATTTGCATAC

10701 CCATGGTGCA CCACGAAAAG ATACCCTCAA GATGCTTGCA CTCCCTCTGT

10751 GCGCGCATTT CTGCACTGTT TTAGAGCATG ATGCCTCTTA CACGCATCTG

10801 TGTGCATAAA CTACATATAG GGAGTGCGTA CCACGCAGGC ATCCAACAAC

10851 CATAAGTGTG TTAAGTGTTA GTTCTCCCTG CGAGGTTCGA AGCGGAAGTC

10901 ACGAATATAC TCGGGTTTCT CTTCAAAGCG CATAAATCTT TCGCCTTTTA

10951 CTAAAGATTT CCGTGGAGAG AAAGTTGTGA GTTTTTATTC AATTTTTTGA

11001 GGCCTCTTAT TTCCTGAGGC TACATTTTTA AGTATTAAAA GTTAGGCAAC

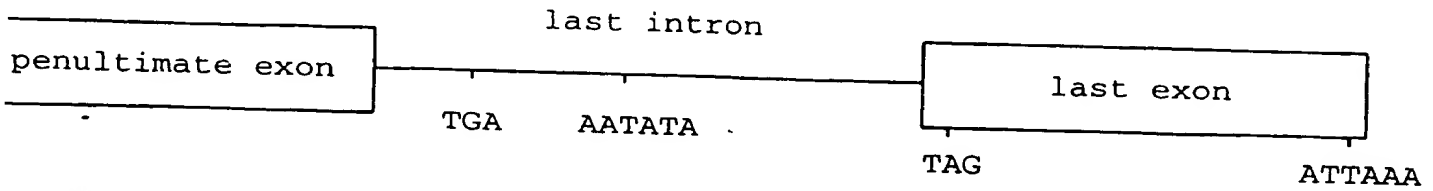11051 TACAAAAAAA AAAAAAAA

Fig 3 (forts)

```
  1 .................................MTRSPPLRELP....... 11
                                     . |...||   :
  1 MASAGNAAEPQDRGGGGSGCIGAPGRPAGGGRRRRTGGLRRAAAPDRDYL 50

 12 ..PSYTPPARTAAPQI...LAGSLKAPLWLRAYFQGLLFSLGCGIQRHCG 56
      |||..:|   |  .||     |.: |||||||| || |||.||| ||:.||
 51 HRPSYCDAA.FALEQISKGKATGRKAPLWLRAKFQRLLFKLGCYIQKNCG 99

 57 KVLFLGLLAFGALALGLRMAIIETNLEQLWVEVGSRVSQELHYTKEKLGE 106
    |.|.:||| |||:|:||:  | :|||:|:|||||||:|||.||:||::|:||
100 KFLVVGLLIFGAFAVGLKAANLETNVEELWVEVGGRVSRELNYTRQKIGE 149

107 EAAYTSQMLIQTARQEGENILTPEALGLHLQAALTASKVQVSLYGKSWDL 156
    ||  ::.:||:||||||  |.:||.||| ||:.|| ||:||.  :.|.:|
150 EAMFNPQLMIQTPKEEGANVLTTEALLQHLDSALQASRVHVYMYNRQWKL 199

157 NKICYKSGVPLIENGMIERMIEKLFPCVILTPLDCFWEGAKLQGGSAYLP 206
    :..:||||||  :.|.| ::..:|| |:||:|:|||||||||||||:|.|||
200 EHLCYKSGELITETGYMDQIIEYLYPCLIITPLDCFWEGAKLQSGTAYLL 249

207 GRPDIQWTNLDPEQLLEELGPFA.SLEGFRELLDKAQVGQAYVGRPCLHP 255
    |:|.:.|||:|| ::|||| .:.  :::: |:|:||:||::|::|||||:|
250 GKPPLRWTNFDPLEFLEELKKINYQVDSWEEMLNKAEVGHGYMDRPCLNP 299

256 DDLHCPPSAPNHHSRQAPNVAHELSGGCHGFSHKFMHWQEELLLGGMARD 305
    .|  .||:.|||.:|  .:  ::|  |.|||||:|:|:||||||||||::||  .::
300 ADPDCPATAPNKNSTKPLDMALVLNGGCHGLSRKYMHWQEELIVGGTVKN 349

306 PQGELLRAEALQSTFLLMSPRQLYEHFRGDYQTHDIGWSEEQASTVLQAW 355
    . |.|:.|.||| |||.|:|:|:||||:|  .|.|.|.:.:|..:|:||
350 STGKLVSAHALQTMFQLMTPKQMYEHFKGYEYVSHINWNEDKAAAILEAW 399

356 QRRFVQLAQEALPENASQQIHAFSSTTLDDILHAFSEVSAARVVGGYLLM 405
    || :|::..::.::::|..|.: .|..||||||||..||:||. ||.:|||||
400 QRTYVEVVHQSVAQNSTQKVLSFTTTTLDDILKSFSDVSVIKVASGYLLM 449

406 LAYACVTMLRWDCAQSQGSVGLAGVLLVALAVASGLGLCALLGITFNAAT 455
    |||||:||||||||..|||.||||||||||||.||.||||.|:||.|||||
450 LAYACLTMLRWDCSKSQGAVGLAGVLLVALSVAAGLGLCSLIGISFNAAT 499

456 TQVLPFLALGIGVDDVFLLAHAFTEALPG..TPLQERMGECLQRTGTSVV 503
    ||||||||||:|||||||||||||:.|.. .|:::| ||||.|||.||.
500 TQVLPFLALGVGVDDVFLLAHAFSETGQNKRIPFEDRTGECLKRTGASVA 549

504 LTSINNMAAFLMAALVPIPALRAFSLQAAIVVGCTFVAVMLVFPAILSLD 553
    ||||.|:..||:.||||||||||||||||:||. .|. |:|:||||||:|
550 LTSISNVTAFFMAALIPIPALRAFSLQAAVVVVFNFAMVLLIFPAILSMD 599

554 LRRRHCQRLDVLCCFSSPCSAQVIQILPQELGDGT............VPVG 592
    |||. .|||||::|||.|||  ..|||: ||...|.
600 LYRREDRRLDIFCCFTSPCVSRVIQVEPQAYTDTHDNTRYSPPPPYSSHS 649

593 IAH.....LTATVQAFTHCEASSQHVVTILPPQAHL....VPPPSDPLGS 633
    :||     :.|||  |.::: .:  |.|.|.:  |... |.|::
650 FAHETQITMQSTVQLRTEYDPHTHVYYTTAEPRSEISVQPVTVTQDTLSC 699
```

Fig. 4A

Genomic

penultimate exon	last intron		last exon

```
                    TGA      AATATA  .
                                        TAG                    ATTAAA
```

```
            GlyLeuArgTrpGlyAlaSerSerSerLeuProGlnSerPheAlaArgValThr
            GGGCTTAGgtgggggggcatcctcctccctgccccagagctttgccagagtgact
            -----C-------acaggc--c--a------------t--------------

            ThrSerMetThrValAlaIleHisProProProLeuProGlyAlaTyrIleHis
            acctccatgaccgtggccatccacccaccccccctgcctggtgcctacatccat
            -----------t------c----------a--t-------a--a---g----c

            ProAlaProAspGluProProTrpSerProAlaAlaThrSerSerGlyGlnLeu
            ccagcccctgatgagccccccttggtcccctgctgccactagctctggcaacctc
            ------t-c--g------a-a-a-
                                    *

            SerSerArgGlyProGlyProAlaThrGly *              gProAsnAsnIle *
            agttccaggggaccaggtccagccactgggtga..........agGCCCGAGGAGATCTAG
```
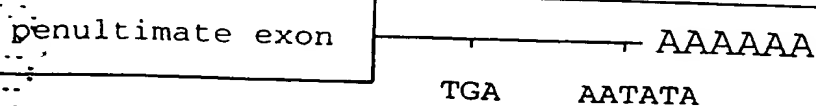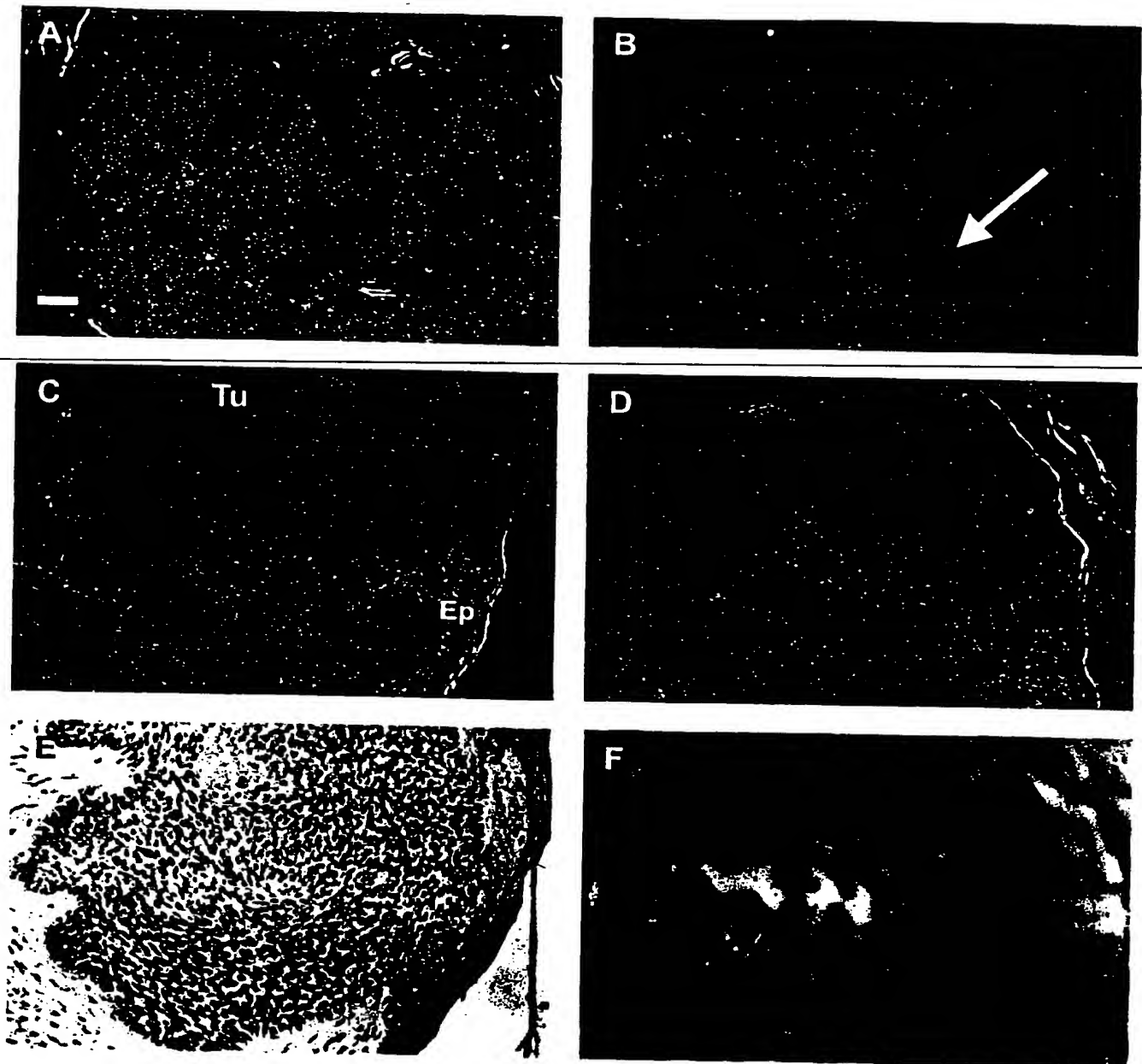
Parotid gland
Colon

penultimate exon			last exon	AAAAAA

```
                                        TAG                    ATTAAA
```

Fetal brain

penultimate exon	

```
                    TGA      AATATA       AAAAAA
```

Fig. 4B

Fig. 4C

Fig. 5